AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 99:369–377 (1996)

A Monte Carlo Study of the Inferential Properties of Three Methods of Shape Comparison

W. MARK COWARD AND DEIRDRE McCONATHY
Department of Biomedical Visualization, University of Illinois at Chicago, Chicago, Illinois 60680 (W.M.C., D.M.); and SYSTAT, Inc., Evanston, Illinois 60201 (W.M.C.)

KEY WORDS Procrustes methods, Superimposition, Simulation, Coordinate free approach, Shape analysis

ABSTRACT Three inferential morphometric methods, Euclidean distance matrix analysis (EDMA), Bookstein's edge-matching method (EMM), and the Procrustes method, were applied to facial landmark data. A Monte Carlo simulation was conducted with three sample sizes, ranging from $n = 10$ to 50, to assess type I error rates and the power of the tests to detect group differences for two- and three-dimensional representations of forms. Type I error rates for EMM were at or below nominal levels in both two and three dimensions. Procrustes in 2D and EDMA in 2D and 3D produced inflated type I error rates in all conditions, but approached acceptable levels with moderate cell sizes. Procrustes maintained error rates below the nominal levels in 2D. The power of EMM was high compared with the other methods in both 2D and 3D, but, conflicting EMM decisions were provided depending on which pair (2D) or triad (3D) of landmarks were selected as reference points. EDMA and Procrustes were more powerful in 2D data than for 3D data. Interpretation of these results must take into account that the data used in this simulation were selected because they represent real data that might have been collected during a study or experiment. These data had characteristics which violated assumptions central to the methods here with unequal variances about landmarks, correlated errors, and correlated landmark locations; therefore these results may not generalize to all conditions, such as cases with no violations of assumptions. This simulation demonstrates, however, limitations of each procedure that should be considered when making inferences about shape comparisons.

A common problem one faces when analyzing biological data is the assessment of similarity between pairs or groups of objects. Methods that produce qualitative results are available (see Richtsmeier et al., 1992, and Bookstein, 1991 for a review of methods), but such methods are insufficient because they do not provide statistical tests of group differences. Inferences to group populations are lacking in qualitative procedures. As a result, using purely qualitative procedures, two individuals may draw different conclusions from the same results. The need for probabilistic judgments has led to the development of quantitative approaches to shape comparison (Bookstein, 1991; Lele and Richtsmeier, 1991; Goodall and Mardia, 1993) that allow inferences to be drawn about the populations from which samples are taken. Because these procedures have emerged recently in the literature, questions remain unanswered regarding the merits and liabilities of these procedures.

Received June 14, 1993; accepted July 26, 1995.

Address reprint requests to W. Mark Coward, Department of Biomedical Visualization (M/C 527), University of Illinois at Chicago, 1919 West Taylor Street, Chicago, IL 60680-6998.

Morphometric tools for inferential testing should have certain characteristics. First, tests must control the type I error rate to alpha, the level of significance chosen for analysis. Increased type I error rates too frequently lead researchers into thinking differences exist when they do not. Second, morphometric methods must be applicable to both two- and three-dimensional data. To date, the majority of morphometric analyses have been restricted to two dimensions, but the increasing availability of tools for data collection in three dimensions challenges and obligates morphometricians to provide suitable statistical tools for the analyses of 3D data. Third, tests ought not to rely on statistical assumptions that are not tenable for biological data. For example, classical statistical assumptions based on the Gaussian perturbation model, such as spherical error variance around landmarks, may not apply to biological forms. If, however, assumptions are made, the test must be robust to violations of those assumptions. Fourth, the power of a test (1 − type II error rate) ought to be sufficient to detect biologically important differences.

The three methods predominantly used for inferential analysis of form are Euclidean distance matrix analysis (EDMA), Bookstein's edge-matching method (EMM), and Procrustes analysis. The degree to which each of these methods conform to all the requirements of morphometric analyses identified above is unclear; hence the current study. A Monte Carlo simulation was conducted to assess the performance of each of these methods with real two- and three-dimensional data. Though the behavior of these tests is of interest in numerous conditions, this study focuses on the two-group problem, where data from two populations are collected with the intent to compare form similarity.

MORPHOMETRIC METHODS

Euclidean distance matrix analysis

In Euclidean distance matrix analysis (Lele and Richtsmeier, 1991), the statistical test of form similarity compares mean forms without attempt to distinguish between size and shape differences. With N landmarks in K dimensions, the comparison of two groups of forms starts with the mean form matrix of each group, an N-by-N symmetric matrix of Euclidean distances computed in K dimensions. According to Lele and Richtsmeier (1991), this matrix can be computed one of two ways. The first alternative is to compute the mean location of each landmark by applying generalized Procrustes analysis (Gower, 1975). The resulting Euclidean distances between mean landmark locations then serve as the distances for the mean form matrix. The second alternative is to compute the matrix of Euclidean distances between all possible pairs of landmarks for each observation and use the arithmetic average of each pair of landmarks to establish the mean form matrix. Lele and Richtsmeier (1991) point out the latter method is biased, but is essentially consistent under a general set of circumstances. To compare two groups, one mean form matrix is divided by the other, resulting in the form difference matrix. A matrix of ones indicates that the forms are precisely the same in terms of size and shape. A matrix of constants other than one is produced when forms differ by size only. For example, a form difference matrix composed entirely of numbers close to two indicates a scale difference, but not a shape difference. Variable numbers in the form difference matrix suggest that the groups differ by size and shape. The test statistic T (the ratio of the largest element to the smallest element of the form difference matrix) is computed to test if the form difference matrix is one of constants. Since T has no defined distribution for comparison, a bootstrap procedure is used to estimate its distribution under the null hypothesis of no group differences. The obtained T statistic is then compared to a predefined cumulative percentile of the bootstrap distribution to arrive at a decision rule.

The specific bootstrap procedure described in Lele and Richtsmeier (1991) follows, according to their notation and description (p. 419).

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be the two samples. Let $Z = (Z_1, Z_2, \ldots, Z_{n+m})$, denote the mixed sample made up of X and Y.

Step 1. Select Z_i^*, i = 1, 2, . . . , n + m from Z randomly and with replacement.
Step 2. Split the bootstrap sample $Z^* = (Z_1^*, Z_2^*, \ldots, Z_{n+m}^*)$ in two groups $Z_1^*, \ldots, Z_n^*$ and $Z_{n+1}^*, \ldots, Z_{n+m}^*$ corresponding to the size of the original samples X and Y.
Step 3. Calculate T* for these two "samples", using the average form obtained by (methods described above).
Step 4. Repeat steps 1–3 B times where B is large (approximately 100).

Lele describes the alternative approach (personal communication, 1993) as follows:

Let Population 1 be defined as $X_1, X_2, \ldots, X_n$ and Population 2 be defined as $Y_1, Y_2, \ldots, Y_m$. Let Population 1 be the base population.
Step 1. Generate $X_1^*, X_2^*, \ldots, X_n^*$ and $Y_1^*, Y_2^*, \ldots, Y_n^*$ from $X_1, X_2, \ldots, X_n$ with replacement.
Step 2. Calculate T* based on this bootstrap sample.
Step 3. Repeat steps 1–3 B times where B is large (approximately 100).

The EDMA test described in Lele and Richtsmeier (1991) assumes the two groups under consideration have the same variance/covariance matrix between landmarks, or if the groups differ in scale, a variance/covariance matrix that differs only by the scaling factor. EDMA does not require assumptions relating to equal variances or Gaussian distributions about landmarks.

Edge-matching method (Bookstein's shape coordinates)

The *edge-matching method* (EMM), otherwise known as the shape coordinate method, described by Bookstein (1986) is an approach that distinguishes between size and shape and allows independent assessment of either attribute. The centroid of a form is the arithmetic mean location of each landmark. The size of an object is defined as the simple sum of squared distances between each landmark and the centroid.

To describe shape, EMM relies on the selection of baseline landmarks and subsequent scaling and rotation of other landmarks with respect to these baseline points. This transformation produces *shape coordinates* for the configuration of points based on the baseline pair. In two dimensions, the pair of baseline points located at 0,0 and 1,0 define shape coordinates for all other landmarks in the new shape space through a simple geometric transformation. In three dimensions landmarks are standardized relative to three points (Goodall and Mardia, 1993) where the shape coordinates of the triangle is a pair of numbers representing the degrees of geometric freedom for shape after scale, translation, and rotation have been removed. These shape coordinates are used in two ways: (1) to assess location relative to meaningful reference points that allow substantive interpretation of shape change or difference (e.g., the gnathion moves down relative to the sinister and dexter endocanthion); and (2) to compare the location of a landmark of interest for two or more groups (e.g., the location of the gnathion is different for males and females).

Different choices for baseline points using EMM will produce different shape coordinates, but the overall description of shape should be consistent no matter which points are selected. For example, in a comparison between two groups of forms for three landmarks, A, B, and C, the statistical test of the location of landmark C with respect to landmarks A and B should match exactly the results comparing the location of A using B and C as baseline points; this is not, however, the case. Bookstein (1991) states that the effects of baseline point selection are mild for small shape differences and provides arguments as to why the differences are negligible.

Relying on multivariate analysis of variance (MANOVA) of the shape coordinates for a decision rule, the assumptions involved in EMM are those of MANOVA. The shape coordinates must be normally distributed and groups must share a common covariance matrix. The central limit theorem states that the assumption of a multivariate normal distribution can be relaxed with large sample sizes.

Procrustes analysis

An extension of Procrustes superimposition (Gower, 1975) allows one to test the

quality of shape of two or more groups of objects (Goodall, 1991). Procrustes analysis entails translating, rotating, and scaling an object onto a target object minimizing some function, usually the sum of squares error between the two objects. Using Goodall's two-sample test for shape data (Goodall, 1991), two groups of objects are compared using the sum of squares residual from superimposing one mean form onto the other and comparing that with the within-groups residual sum of squares. Using the notation from Goodall (1991, page 290), the F test is constructed as follows:

With N landmarks in K dimensions
Shape dimension
$m = N^*K - \frac{1}{2}K(K + 1) - 1$
Sample sizes of group x and y of L_x and L_y
Residual sum of squares between the two forms G*
Procrustes sums of squares for x after superimposition G(X′)
Procrustes sums of squares for y G(Y)

$$F_{m,(L_x + L_y - 2)m} = \frac{L_x + L_y - 2}{L_x^{-1} + L_y^{-1}} \frac{G^*}{G(X') + G(Y).}$$

This test makes three assumptions: Gaussian distribution of landmarks in the K dimensions, landmarks are uncorrelated, and errors are uncorrelated within forms.

METHODS

Type I and type II error rates for tests of group differences, either shape or size, between two groups were assessed for EDMA, EMM, and Procrustes analysis. To establish a population from which data were sampled, 13 adult female (living) heads were scanned with a Cyberware Laboratory 3D Digitizer (model 5020/PS). This scanner provided 256,000 data points in three dimensions for each subject. The scanner produced a rendered image from which four facial landmarks in x, y, and z space were selected for each subject. An expert located four landmarks for each of the women: sinister and dexter exocanthion, stomion, and gnathion using the program LEGO (Neumann, 1992). Abbrevations for these landmarks appear in Table 1. Once locations were established it was necessary to estimate the covariance matrix of the landmarks.

TABLE 1. Abbreviations

EXS	Sinister exocanthion
EXD	Dexter exocanthion
GN	Gnathion
STO	Stomion

Three methods could be used to estimate the variance/covariance matrix of landmarks. For tests described by Goodall (1991), the matrix is assumed to be in the form of an identity matrix, where neither landmarks nor errors are correlated. As a second alternative, one can relax the assumptions somewhat and allow correlations between landmarks, but not errors. It is our contention that neither restraint on the covariance matrix is plausible with real morphometric data. Our experience suggests that landmarks on forms are highly correlated, and differences from mean landmark locations and the landmarks themselves are almost never independent. This led us to use a third method to estimate the variance/covariance matrix of landmarks.

The forms were translated and rotated (with ordinary Procrustes analysis) with respect to a common target object to align the forms. Once aligned, the mean landmark positions and intercorrelations between points defined the population. We believe that this approach is appropriate in biological contexts. This method does not make untested assumptions about the correlations between landmarks, such as the absence of correlation between points, and does not impose an arbitrary structure upon the data in determining covariance. Nature defines the parameters of the populations from which biological forms are taken, not statisticians. Our intent is to use plausible data and to apply techniques to the data that simulate experimental conditions routinely encountered by morphometricians. While other procedures are preferred by some for population covariance estimate, our selection for this problem is appropriate. We want only to estimate a plausible estimate of the population characteristics, which this method produces.

The initial population defined by locations

TABLE 2. Group populations (rounded for display)

	Landmark	x Mean (SD)	y Mean (SD)	z Mean (SD)
Control group	Sinister exocanthion	4.464 (0.284)	3.784 (0.287)	1.103 (0.119)
	Dexter exocanthion	−4.199 (0.253)	4.126 (0.322)	1.668 (0.111)
	Stomion	−0.068 (0.092)	−2.255 (0.205)	−1.361 (0.224)
	Gnathion	−0.197 (0.079)	−5.655 (0.421)	1.410 (0.157)
Comparison group 2D (differences only)	Stomion			−1.236
	Gnathion			1.385
Comparison group 3D (differences only)	Sinister exocanthion			1.228
	Gnathion			1.385

and the variance/covariance matrix of the four facial landmarks provided the basis for manipulation of data points to create a differing sample of landmarks for comparison. Within each study, the second population was identical to the first, except one or more points were moved by a small uniform amount. Both populations had the following characteristics: (1) unequal (nonspherical) variance within landmarks; (2) unequal variance between landmarks; (3) correlated landmark locations; (4) correlated errors within forms; and (5) equal variance; covariance matrices. Means and standard deviations for the two groups are shown in Table 2.

Random numbers were computed with a random number generator described by Wichmann and Hill (1982). Data were generated with population means and correlations as described by Wilkinson (1990). Linear model statistics were computed with a modified version of SYSTAT's MGLH version 5.03 program (Wilkinson, 1990). EDMA analyses were conducted with the computer program SHAPE (Lele and Richtsmeier, 1991) that uses the mean form matrix derived by averaging Euclidean distances between landmarks. The estimated T distribution was derived from the alternative method described above. Procrustes analysis was performed with software written by W.M.C. for this task using SYSTAT's statistical library and probability routines. All statistical tests were declared significant with an obtained alpha level less than 0.05.

Type I error rate

The type I error rate, the number of times a test falsely rejected the null hypothesis of equality, was assessed for each procedure. Ten thousand replications at three sample sizes (n = 10, 30, 50) of two groups drawn from the female form population described above were analyzed with each statistical procedure. Since all tests were of groups from the same population, each resulting decision rule, if correct, would not reject the null hypothesis of equality.

Two-dimensional analyses involved the stomion, dexter exocanthion, and gnathion in the x and z dimensions. Three-dimensional analyses included all landmarks (sinister and dexter exocanthion, stomion, and gnathion) described above.

Type II error rate

The type II error rate (proportion of the time a test failed to reject the null hypothesis of equality when the data were from different populations) for each method was evaluated at three sample sizes (n = 10, 30, 50). Ten thousand analyses in each sample size were performed with the landmark selections described above.

RESULTS

Type I error rates

Table 3 shows the type I error rate with two-dimensional data. The columns "All EMM" and "Majority EMM" are the percentage of instances where each of the baseline pairs and a majority of pairs (greater than one-half) rejected the null hypothesis, respectively. With each of the different baseline reference points EMM consistently held the type I error rate approximately at or under the nominal level. Both comparisons, "All EMM" and "Majority EMM" tests, considered simultaneously, maintained an error

TABLE 3. Two-dimensional type I error rates

Sample size	Procrustes	EDMA	EMM				
			STO-EXD	STO-GN	EXD-GN	All EMM	Majority EMM
10	4.20	11.06	5.07	3.27	4.91	0.90	4.63
30	4.11	7.15	5.08	3.97	4.93	0.55	4.41
50	4.71	6.70	4.93	3.74	4.96	0.47	4.42

TABLE 4. Three-dimensional type I error rates

Sample size	Procrustes	EDMA	EMM					
			EXS-EXD-STO	EXS-EXD-GN	EXD-STO-GN	EXS-STO-GN	All EMM	Majority EMM
10	10.95	9.51	4.65	4.87	4.56	4.57	1.43	3.00
30	10.40	8.12	5.39	5.23	5.30	5.31	1.75	3.44
50	10.42	6.34	5.17	5.16	4.82	5.18	1.66	3.38

TABLE 5. Percent of correct rejections with 2D data

Sample size	Procrustes	EDMA	EMM					
			STO-EXD	STO-GN	EXD-GN	% Conflicting results	All EMM	Majority EMM
10	17.72	24.09	29.93	10.50	30.18	27.21	7.07	29.26
30	58.48	42.86	79.54	21.73	79.54	61.63	20.21	78.76
50	84.90	61.48	95.81	29.66	95.64	67.15	29.28	95.40

rate below alpha. EDMA demonstrated inflated type I error rates of several percent and approached alpha with larger sample sizes. It appears from the table that EDMA converges to alpha with sample sizes somewhat above 50. Procrustes was consistently below the nominal alpha, ranging between 4.2 and 4.71%.

Table 4 shows the type I error rate for the three-dimensional analyses. The general pattern of the two-dimensional type I error rate was approximately replicated in three dimensions except for Procrustes. EMM consistently kept levels approximately at the nominal level or below with simultaneous tests keeping the errors below alpha. Procrustes produced type I error rates between 10 and 13%, appearing to converge to alpha as sample size increased. EDMA exhibited inflated error rates clearly approaching alpha with larger sample sizes.

Type II error rates

Table 5 shows the power of each test with the two-dimensional data. Power (1 − type II error rate) is reported rather than the type II error rate since it is easier to read; higher percentages are "better." As with type I error rates, the power was computed for EMM using all pairs of points as baseline. Since three forms of EMM were applied to the same data, it is possible to have all three tests in agreement or have conflicting results. The column "% conflicting results" is the percent of instances when all EMM tests were not in agreement (not all rejecting or all accepting the null hypothesis). A test based on all pairs and a majority of pairs was computed as described above.

In two dimensions, each test improved sensitivity to true differences as sample size increased. EMM baseline pairs STO-EXD and EXD-GN were most powerful. These tests also tended to reject the null hypothesis consistently, given the agreement of the two reflected by the high majority rejection rate of the three pairs. EDMA ranked second with smaller sample sizes, and Procrustes analysis ranked second with larger sample sizes. The STO-GN EMM baseline pair was the

TABLE 6. Percent of correct rejections with 3D data

Sample size	Procrustes	EDMA	EMM						
			EXS-EXD-STO	EXS-EXD-GN	EXD-STO-GN	EXS-STO-GN	% Conflicting results	All EMM	Majority EMM
10	13.16	10.63	7.52	8.33	7.43	7.35	11.80	2.82	5.00
30	12.27	8.59	15.94	17.84	15.35	14.54	20.53	6.84	11.81
50	10.97	7.69	25.65	29.14	24.84	22.29	27.89	12.37	20.32

least powerful in all instances. The three EMM baseline pairs produced conflicting results between 27.21% and 67.15% of the time, largely due to the STO-GN pair being less powerful than the other baseline pairs. Each EMM pair correctly rejected the null hypothesis at a rate very close to the STO-GN pair, illustrating that that pair was the upper bound of the simultaneous hit rate.

In three dimensions, EMM increased power with larger sample sizes (Table 6). The rate of conflict between the EMM baseline pairs invariably increased with sample size, ranging from 11.80% to 27.89%. Given the large range of power of particular pairs, each pair correctly rejected the null hypothesis relatively infrequently between 2.82% and 12.37%. The majority of the tests followed closely behind the general pattern of rejections of the individual tests, showing general agreement between most of the tests with most of the samples. EDMA and Procrustes analysis tended to decrease in power with larger sample sizes in both studies, a result that is difficult to explain. In terms of relative power in three dimensions, individual EMM tests produced the highest rejection rate, followed by Procrustes and EDMA.

DISCUSSION

These tests were applied to a specific case where certain assumptions were violated and relatively few landmarks were chosen for analysis. These results are not necessarily generalizable to situations at large where the assumptions are violated, the assumptions are tenable, or there are a larger number of landmarks. The results do reveal, however, how the tests may perform with real data that violate assumptions with only a few landmarks.

Each test revealed undesirable characteristics in this simulation. Procrustes analysis and EDMA had inflated type I error rates and low or unusual power characteristics. EMM produced a relatively large number of conflicting results depending on which landmarks were chosen as baseline points.

While the two-dimensional Procrustes results controlled type I error rate and were relatively powerful, the three-dimensional behavior was less desirable, with an increased type I error rate and an inverse relationship between sample size and power. The three-dimensional qualities can partly be explained by Slice (1993), illustrating that with more landmarks the estimation of the rotation, translation, and scale parameters improves, producing a more powerful test. We found that unequal variances may also have affected the power characteristics of the test found here. In an informal extension of this study, we subjected the populations precisely as described here, except with uniform variances, to the same tests of type I and II error rates. Preliminary findings suggest that the unequal variances contribute to the lower power and inflated error rates.

EDMA results tend to be too liberal with smaller sample sizes, at least with sample sizes of 50 and less. In both two and three dimensions, EDMA approached, but did not achieve, the nominal error rate. EDMA demonstrated behavior similar to the Procrustes test in three dimensions in terms of decreasing power with larger sample sizes. Although EDMA does not make the assumption of equal variances, the findings from the informal extension of this study suggest that the unequal variances may contribute to the inflated error rates with small samples as well as the odd power characteristics in three dimensions.

EMM consistently maintained type I error rates about or below the nominal level in

TABLE 7. Two-dimensional type I error rates (pilot study results)

Sample size	Procrustes	EDMA	EMM		
			EXS-EXD	EXS-GN	EXD-GN
10	7.99	8.65	4.49	4.64	4.62
30	8.73	6.47	5.14	5.25	5.19
50	7.61	5.39	4.66	4.66	4.94

TABLE 8. Three-dimensional type I error rates (pilot study results)

Sample size	Procrustes	EDMA	EMM			
			EXS-EXD-STO	EXS-EXD-GN	EXD-STO-GN	EXS-STO-GN
10	12.97	9.88	4.71	4.87	5.03	3.95
30	11.47	6.23	4.68	4.77	5.07	4.38
50	11.96	5.54	4.65	4.68	4.82	4.64

TABLE 9. Proportion of correct rejections with 2D data (pilot study results)

Sample size	Procrustes	EDMA	EMM				
			EXS-EXD	EXS-GN	EXD-GN	% Conflicting results	All EMM
10	8.58	68.01	46.75	48.19	43.76	16.62	37.82
30	8.96	97.45	94.53	95.18	92.16	6.23	90.58
50	9.57	99.88	99.69	99.75	99.28	.70	99.15

TABLE 10. Proportion of correct rejections with 3D data (pilot study results)

Sample size	Procrustes	EDMA	EMM					
			EXS-EXD-STO	EXS-EXD-GN	EXD-STO-GN	EXS-STO-GN	% Conflicting results	All EMM
10	17.02	10.69	47.62	35.86	27.42	16.45	50.63	5.16
30	28.48	9.39	96.39	88.78	77.84	35.53	70.28	27.12
50	45.80	9.95	99.79	98.87	95.59	53.72	48.79	51.10

both two and three dimensions. For all but one baseline pair, the power of the EMM test was similar to or exceeded other tests except with small sample sizes. The most significant problem, though, is that often conflicting results were obtained both in 2D and 3D. Conflicting decision rules as high as 29.28% were found for results from different baseline points. This is a troublesome finding. What should one do if the location of C (relative to A and B) differs by group, but the location of A (relative to B and C) does not? Which is the "correct" decision? Do the groups differ in terms of shape?

Formal research is required in several areas to overcome the difficulties described here. First, a simulation study of the individual effect of unequal variances, correlated errors, and correlated landmark locations will shed light on which of these factors most effects type I error rate and power. From this study, it is unclear in what proportion these factors influenced the tests. Second, we need statistical tests void of untenable assumptions, possibly including a model of correlated errors, correlated landmarks, unequal variances about landmarks, and nonspherical variance about each landmark. What is clear from this study, however, is that limitations of each procedure must be considered when making inferences regarding shape comparisons.

A NOTE REGARDING PRELIMINARY MONTE CARLO STUDIES

We carried out a pilot study prior to the work described here. In that study population parameters were estimated in the identical manner as described above except using four, rather than 13, heads to estimate the population parameters. Hence, the covariance matrix of the landmark locations in three dimensions was singular providing the dimensional data plots showing landmarks as "disks" rather than points that form multivariate normal distributions. This population may not represent plausible biological variability because the covariance matrix is not of full rank.

Notwithstanding the limits in interpretation of results based on a singular population matrix, the comparative results may be of interest to the reader because the groups indeed did differ. Tables 7–10 show the analogs of Tables 3–6, respectively, for the pilot data. The statistics pertaining to the majority decision rule of EMM were not calculated in the pilot work and do not appear in these tables.

The general pattern of small sample type I error rates, inflated with Procrustes and EDMA and generally conservative EMM results (Tables 7, 8), replicated across both studies. The statistical power (Tables 9, 10), on the other hand, differed markedly from the study described above. With the two-dimensional data EDMA clearly outperformed the other tests, but at the expense of elevated type I errors. With three-dimensional data the power of the EDMA remained somewhat constant, whereas Procrustes and EMM increased with power as sample sizes increased.

The differences between the pilot and subsequent work can be attributed to at least two causes. First, one may outright dismiss the pilot simulation data because the population may not reflect that of true biological variability, and the results may be an artifact of the population characteristics. Second, one can attribute the differences to the particular population chosen for analysis. We here prefer the latter attribution to performance differences because of the results of unpublished work subsequent to that published in this paper. Though more work is needed to get a better understanding of the type I and type II error rates of these tests, we believe that the differences in power characteristics shown in these two simulations would prevail with other populations based on full rank covariance matrices.

LITERATURE CITED

Bookstein F (1986) Size and shape spaces for landmark data in two dimensions. Stat. Sci. *1*:181–242.

Bookstein F (1991) Morphometric Tools for Landmark Data. Cambridge: Cambridge University Press.

Goodall C (1991) Procrustes methods in the statistical analysis of shape. J. R. Stat. Soc. B. *53*:285–339.

Goodall CR, and Mardia KV (1993) Multivariate aspects of shape theory. Ann. Stat. *21*:848–866.

Gower J (1975) Generalized Procrustes analysis. Psychometrika *40*:33–50.

Lele S (1991) Some comments on coordinate-free and scale-invariant methods in morphometrics. Am. J. Phys. Anthropol. *85*:407–417.

Lele S, and Richtsmeier J (1991) Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. Am. J. Phys. Anthropol. *86*:415–427.

Newmann PF (1992) LEGO: A Visualization Package for 3D Laser Scanned Objects. Master's thesis, University of Illinois at Chicago.

Richtsmeier J, Cheverud J, and Lele S (1992) Advances in anthropological morphometrics. Ann. Rev Anthropol. *21*:283–305.

Slice DE (1993) Extensions, Comparisons, and Applications of Superimposition Methods for Morphometric Analysis. PhD dissertataion, Department of Ecology and Evolution, State University of New York at Stony Brook.

Wichman BA, and Hill ID (1982) An efficient and portable pseudo-random number generator. Algorithm AS 183. Appl. Stat. *311*:188–190.

Wilkinson L (1990) SYSTAT: The System for Statistics. Evanston, IL: SYSTAT, Inc.